APPLICATION

FOR

UNITED STATES LETTERS PATENT

TITLE: TRANSCRIPTIONAL REGULATOR

APPLICANT: MICHAEL H. JONES

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL983024483US

November 5, 2003
Date of Deposit

TRANSCRIPTIONAL REGULATOR

This is a continuation-in-part of PCT/JP98/01783, filed April 17, 1999, and claims priority from Japanese Application Nos. 9/116570, filed April 18, 1997, and 9/310027, filed October 24, 1997.

Technical Field

The present invention relates to a novel transcriptional regulator containing a bromodomain and a gene encoding it.

Background Art

The bromodomain is a characteristic motif of proteins found in transcriptional regulators. Proteins having a bromodomain usually contain one or two (Tamkun, J. W. et al., (1992), Nuc. Acids Res., 20:2603), but sometimes as many as five bromodomain motifs (Nicolas, R. H. and Goodwin, G. H. (1996), Gene, 175 (12):233-240). This motif is found in a wide variety of animals. For example, it is identified in yeast (Winston, F. et al., (1987), Genetics, 115:649-656; Laurent, B. C. et al., (1991), Proc. Natl. Acad. Sci. USA, 88:2687-2691), in Drosophila (Digan, M. E. et al., (1986), Dev. Biol., 114:161-169; Tamkun, J. W. et al., (1992), Cell, 68:561-572), and in the genes for transcriptional regulators in mammals (Denis, G. V. and Green, M. R. (1996), Genes and Devel., 10:261-271; Yang, X. J. et al., (1996), Nature, 382:319-324).

All transcriptional regulators containing a bromodomain serve to control signal-dependent transcription in actively proliferating cells (Tamkun, J. W. et al., (1992), Cell, 68:561-572; Haynes, S.R. et al., (1992), Nuc. Acids Res., 20:2603). Due to this feature of these transcriptional regulators, it is suggested that cancer may

develop if the gene for the protein containing a bromodomain is not normally controlled. In fact, several studies have shown that human transcriptional regulators with a bromodomain RING3, p300/CBP, and PCAF may be involved in oncogenesis.

RING3 is a transcriptional regulator highly homologous with the fsh protein that regulates development of Drosophila (Haynes, S. R. et al., (1989), Dev. Biol., 134:246-257). RING3 is a nuclear serine/threonine kinase having autophosphorylating activity. This activity of RING3 correlates with a proliferating state in chronic or acute lymphocytic leukemia. For instance, when Denis and Green collected lymphocytes of peripheral blood from 10 patients suffering from leukemia, kinase activity associated with RING3 was identified in all of the 10 patients but not in normal controls (Denis, G. V. and Green, M. R. (1996), Genes and Develop., 10:261-271). Furthermore, this activity was not detected in the blood cells from patients whose leukemia had remitted by virtue of chemotherapy.

p300 and CBP (CREB binding protein) encode highly similar proteins and are thus often called p300/CBP. p300/CBP is a co-activator for a transcriptional regulator CREB (cAMP responsive element binding protein) (Kwok, RPS et al., (1994), Nature, 370:223-226), and is considered as a key protein for growth regulation. Mutation in p300/CBP has been found in familial or sporadic cancers. Germline mutation of CBP results in Rubinstein-Taybi syndrome, which causes patients to develop various malignant tumors (Petrij, F. et al., (1995), Nature, 376:348-51), while mutation in p300 is found in sporadic colorectal and gastric cancers (Muraoka, M. et al., (1996), Oncogene, 12:1565-1569). Furthermore, CBP is fused with MOZ (Monocytic leukemia Zinc finger protein) in a t (8; 16) (p11; p13) translocation

found in a certain kinds of acute myelocytic leukemia. The fusion protein has histone-acetyltransferase domains derived from both genes (Bannister, A. J. and Kouzarides, T. (1996), Nature, 384:641-643; Orgyzco, V. V. et al., (1996), Cell, 87:953-959; Brownwell, J. E. and Allis, C. D. (1996), Curr. Opin. Genet. Devel., 6:176-184). Since acetylated histone is known to be associated with transcriptionally active chromatin, the fusion protein may be involved in leukemogenesis by way of aberrant histone acetylation (Brownwell, J. E. and Allis, C. D. (1996), Curr. Opin. Genet. Devel., 6:176-184).

p300/CBP is also considered to be associated with cancer since it interacts with known oncogene products. For example, p300/CBP binds to E1A protein (Arany, Z. et al.,(1995), Nature, 374:81-84), one of the early genes of adenovirus. p300 is also a co-activator for transcription factors, c-Myb (Dai, P. et al., (1996), Genes Dev., 10:528-540) and c-Fos (Bannister, A.J. and Kouzarides, T. (1996), Nature, 384:641-643).

PCAF, is considered to inhibit the interaction of E1A with p300/CBP by competing with E1A for binding to p300/CBP (Yang, X.J. et al., (1996), Nature, 382:319-324). PCAF also has histone-acetyltransferase activity.

Thus, it is thought that transcriptional regulators containing a bromodomain are involved in regulation ofcell growth, and that their aberrant regulation may be closely related to various diseases, particularly to cancer. Transcriptional regulators containing a bromodomain have thus recently received much attention as novel targets for specifically treating cancer.

Disclosure of the Invention

The objective of the present invention is to provide a novel transcriptional regulator containing a bromodomain and a gene encoding it, and a method of screening for a candidate compound as a medicament by using them.

As a result of research to achieve the above objective, the inventors successfully isolated several genes, each of which encodes a novel transcriptional regulator containing a bromodomain. The genes were isolated from a human testis cDNA library using primers designed based on EST sequences which had been identified using known bromodomain sequences as probes. In addition, the inventors have found that the structures of the isolated genes resemble one another, thus they constitute a family. The inventors have also found that the isolated genes or proteins encoded by them can be used to screen the candidate compounds for a medicament that controls the activity of the proteins or other factors interacting therewith.

Thus, the present invention relates to novel transcriptional regulators each having a bromodomain and the genes encoding them, and to a method of screening for a candidate compound as a medicament using said proteins or genes, and more specifically relates to:

(1) a transcriptional regulator having a bromodomain, which comprises the amino acid sequence shown in SEQ ID NO:1, 13, 21, 27, or 29, or said sequence wherein one or more amino acids are substituted, deleted, or added;

(2) a transcriptional regulator having a bromodomain, which is encoded by DNA hybridizing with DNA comprising the nucleotide sequence shown in SEQ ID NO:2, 14, 22, 28 or 30;

(3) DNA coding for the transcriptional regulator according to (1) or (2);

(4) a vector comprising the DNA according to (3);

(5) a transformant expressibly retaining the DNA according to (3);

(6) a method for producing the transcriptional regulator according to (1) or (2), which comprises culturing the transformant according to (5);

(7) an antibody binding to the transcriptional regulator according to (1) or (2);

(8) a method of screening a compound having binding activity to the transcriptional regulator according to (1) or (2), wherein the method comprises contacting a sample with the transcriptional regulator according to (1) or (2) and selecting a compound having binding activity to the transcriptional regulator according to (1) or (2);

(9) a compound having binding activity to the transcriptional regulator according to (1) or (2), which can be isolated according to the method of (8);

(10) the compound according to (9), which is naturally occurring; and

(11) DNA specifically hybridizing with DNA comprising the nucleotide sequence shown in SEQ ID NO:2, 14, 22, 28, or 30 and having at least 15 nucleotides.

Here, the term "transcriptional regulator(s)" means protein(s) that control gene expression, and "bromodomain" means an amino acid motif conserved among the transcriptional regulators associated with signal-dependent transcription, wherein said motif is involved in protein-protein interaction.

The present invention relates to novel transcriptional regulators having a bromodomain (BAZ family). The nucleotide sequences of cDNA isolated by the inventors, which belong to BAZ family, are shown in SEQ ID NO:2 (BAZ(BAZ1α)), SEQ ID NO:14 (BAZ2α), SEQ ID NO:22

(BAZ2β), and SEQ ID NO:28 and 30 (BAZ1β). The amino acid sequences of proteins encoded by the cDNA are also shown in SEQ ID NO:1 (BAZ(BAZ1α)), SEQ ID NO:13 (BAZ2α), SEQ ID NO:21 (BAZ2β), and SEQ ID NO:27 and 29 (BAZ1β).

The bromodomain is characteristic of a structural region that is conserved among a group of transcriptional regulators involved in signal-dependent transcription (Tamkun, J. W. et al., (1992), Cell, 68:561-572; Haynes, S. R. et al., (1992), Nuc. Acids Res., 20:2603), and it has been reported that the six mammalian genes, i.e., RING3, p300/CBP, PCAF, BRG1, HRX/ALL-1, and TIF1, which encode transcriptional regulators having a bromodomain, are associated with cancer. That the transcriptional regulators having a bromodomain are commonly associated with cancer suggests that the genes isolated by the inventors are also associated with cancer. Other than a bromodomain motif, the proteins encoded by the genes isolated by the inventors share the characteristic motifs of (1) C4HC3 zinc-finger, which is found in the proteins expressed in a wide range of organisms from yeast to human and is believed to be involved in a protein-protein interaction or nonspecific binding to DNA; (2) leucine zipper, which is present in many transcriptional regulators and is known to contribute to form a dimer with the protein itself or other proteins (Busch, S. J. and Sassone-Corsi, P. (1990), Trends in Genetics, 6:36-40); (3)LXXLL motif, a motif commonly found among many transcriptional co-activators, which is shown to be required for mediation of transcription induced by a nuclear receptor (Torchia, J. et al., (1997), Nature, 387:677-684; Heery, D. M. et al., (1997), Nature, 387:733-736); and (4) nuclear transport signal, which confers the transporting activity into the nucleus on the proteins synthesized in the cytoplasm.

The combination of a bromodomain and C4HC3 zinc finger is known to be associated with several breakpoint genes of leukemia (Tkachuk, D. C. et al., (1992), Cell, 71:691-700; Gu, Y. et al., (1992), Cell, 71:701-798; Miki, T. et al., (1991), Proc. Nat. Acad. Sci., 88:5167-5171; Le Douarin B. et al., (1995), EMBO J., 14:2020-2033; Borrow, J. et al., (1996), Nature Genet., 14:33-41). Accordingly, the genes isolated by the inventors are important candidates for breakpoint genes of cancers.

The transcriptional regulators of the present invention can be prepared as recombinant proteins generated using a recombinant gene technique, or as naturally occurring proteins, according to a method known to one skilled in the art. The recombinant proteins can be prepared using a method such as incorporating DNA encoding a transcriptional regulator of the present invention (e.g., DNA having the nucleotide sequence shown in SEQ ID NO:2, 14, 22, 28, or 30) into a suitable expression vector, which is then introduced into host cells, and purifying the protein obtained from the transformant. The naturally occurring proteins can be prepared using a method such as preparing a column which utilizes an antibody obtained from a small animal immunized with the recombinant protein prepared as above, and subjecting the extract from a tissue or cells in which a transcriptional regulator of the present invention is overexpressed (e.g., testis and cancer cells) to affinity chromatography using said column.

The present invention also relates to transcriptional regulators functionally equivalent to the transcriptional regulators of the present invention having the amino acid sequence shown in SEQ ID NO:1, 13, 21, 27, or 29. A method of introducing mutation into amino acids of a protein to isolate a protein functionally equivalent to a

particular protein is well known to one skilled in the art. Thus, it is well within the art of an ordinarily skilled person to isolate a transcriptional regulator functionally equivalent to the transcriptional regulators of the present invention having the amino acid sequence shown in SEQ ID NO:1, 13, 21, 27, or 29 by appropriately modifying, for example, substituting amino acids without affecting the function of the transcriptional regulator. Mutation in an amino acid of a protein can also occur spontaneously. The transcriptional regulators of the present invention include those having a bromodomain and the amino acid sequence of SEQ ID NO:1, 13, 21, 27, or 29 wherein one or more amino acids are substituted, deleted, or added. Examples of known methods for introducing amino acid mutation into the protein are a site-directed mutagenesis system using PCR (GIBCO-BRL, Gaithersburg, Maryland) and a site-directed mutagenesis using oligonucleotides (Kramer, W. and Fritz, H. J. (1987), Methods in Enzymol., 154:350-367). The number of mutagenized amino acids is usually 50 amino acids or less, preferably 30 amino acids or less, more preferably 10 amino acids or less, and most preferably three amino acids or less.

As another method of isolating a functionally equivalent protein utilizing a hybridization technique (Sambrook, J. et al.,, Molecular Cloning 2nd ed. 9.47-9.58, Cold Spring Harbor Lab. press, 1989) is well known to one skilled in the art. Based on the DNA sequence encoding the transcriptional regulator of the present invention shown in SEQ ID NO:2, 14, 22, 28, or 30, or the fragment thereof, a person with ordinary skill in the art can isolate DNA highly homologous to said DNA sequences using a hybridization technique (Sambrook, J. et al., Molecular Cloning 2nd ed. 9.47-9.58, Cold Spring Harbor Lab. press, 1989) to obtain a

transcriptional regulator functionally equivalent to the transcriptional regulators. The transcriptional regulators of the present invention include those encoded by DNA that hybridizes with DNA comprising the DNA sequence shown in SEQ ID NO:2, 14, 22, 28, or 30, and which contains a bromodomain. The hybridization and washing conditions for isolating DNA encoding a functionally equivalent protein are defined as low stringency: 42°C, 2 X SSC, 0.1% SDS; moderate stringency: 50°C, 2 X SSC, 0.1% SDS; and high stringency: 65°C, 2 X SSC, 0.1% SDS. The transcriptional regulators obtained by the hybridization technique may have amino acid homology of preferably 40% or more, more preferably 60% or more, still more preferably 80% or more, or most preferably 95% or more, with the transcriptional regulators having the amino acid sequence shown in SEQ ID NO:1, 13, 21, 27, or 29. In particular, high homology in the bromodomain sequence is considered significant for the function associated with cancer. Functionally equivalent transcriptional regulators to be isolated may contain, other than a bromodomain, a sequence involved in interaction with another protein (e.g., leucine-zipper or LXXLL motif), a sequence involved in binding to DNA (e.g. zinc finger), or a nuclear transport signal. The presence of the bromodomain in the protein can be identified by searching the bromodomain motif PROSITE database on DNASIS (HITACHI Software engineering).

The present invention also relates to DNA that codes for a transcriptional regulator of the present invention. The DNA of the present invention includes cDNA, genomic DNA, and chemically synthesized DNA, but is not limited thereto as long as it codes for a transcriptional regulator of the present invention. cDNA can be prepared, for example, by designing a primer based on the nucleic acid sequence shown

in SEQ ID NO:2, 14, 22, 28, or 30 and performing plaque PCR (see Affara, N.A. et al., (1994), Genomics, 22:205-210). The genomic DNA can be prepared according to a standard technique using, for example, Qiagen genomic DNA kits (Qiagen, Hilden, Germany). The DNA sequence thus obtained can be determined according to a standard technique using a commercially available dye terminator sequencing kit (Applied Biosystems) and the like. In addition to applying to the production of recombinant proteins as described below, the DNA of the present invention may be applied to gene therapy and the like.

The present invention also relates to a vector into which the DNA of the present invention is inserted. There is no particular limitations to the vector into which the DNA of the present invention is inserted, and various types of vectors, e.g. for expressing the transcriptional regulators of the present invention *in vivo* and for preparing recombinant proteins, may be used for each purpose. Vectors used for expressing the transcriptional regulators of the present invention *in vivo* (in particular, for gene therapy) include the adenovirus vector pAdexLcw and the retrovirus vector pZIPneo. A LacSwitch II expression system (Stratagene; La Jolla, CA) is advantageous when mammalian cells, such as CHO, COS, and NIH3T3 cells, are used. An expression vector is particularly useful for producing a transcriptional regulator of the present invention. Although there is no particular limitation to the expression vectors, the following vectors are preferred: pREP4 (Qiagen, Hilden, Germany) when *E. coli* is used; SP-Q01 (Stratagene, La Jolla, CA) when yeast is used; and BAC-to-BAC baculovirus expression system (GIBCO-BRL, Gaithersburg, Maryland) when insect cells are used. The DNA of the

present invention can be inserted into vectors using a standard method.

The present invention also relates to a transformant expressibly retaining the DNA of the present invention. The transformants of the present invention include one harboring the above-described vector into which the DNA of the present invention is inserted and one having the DNA of the present invention integrated into its genome. The DNA of the present invention can be retained in the transformant in any form as long as the transformant expressibly retains the DNA of the present invention. There is no limitation to host cells into which a vector of the present invention is introduced. If the cells are used to express a transcriptional regulator of the present invention *in vivo*, desired cells may be used as target cells. Cells such as *E. coli*, yeast cells, animal cells, and insect cells can be used for producing the transcriptional regulators of the present invention. The vector can be introduced into the cells by methods such as electroporation and heat shock. Recombinant proteins can be isolated and purified from the transformants generated for producing the said proteins according to a standard method.

The present invention also relates to antibodies that bind to the transcriptional regulators of the present invention. The antibodies of the present invention include, but are not limited to, polyclonal and monoclonal antibodies. Also included are antisera obtained by immunizing an animal such as a rabbit with a transcriptional regulator of the present invention, any class of polyclonal or monoclonal antibodies, humanized antibodies generated by gene recombination, and human antibodies. The antibodies of the present invention can be prepared according to the following method. For polyclonal antibodies, antisera can

be obtained by immunizing a small animal, such as a rabbit, with a transcriptional regulator of the present invention, then recovering the fractions that only recognize the transcriptional regulator of the present invention through an affinity column coupled with the transcriptional regulator of the present invention. Immunoglobulin G or M can be prepared by purifying the fractions through a Protein A or G column. For monoclonal antibodies, a small animal, such as a mouse, is immunized with a transcriptional regulator of the invention, the spleen is removed from the mouse and homogenized into cells, the cells are fused with myeloma cells from a mouse using a reagent such as polyethylene glycol, and clones that produce antibodies against the transcriptional regulator of the invention are selected from the resulting fused cells (hybridoma). The hybridoma obtained is then transplanted into the abdominal cavity of a mouse, and the ascites are recovered from the mouse. The obtained monoclonal antibodies can then be prepared by purifying , for example, by ammonium sulfate precipitation through a Protein A or G column, by DEAE ion exchanging chromatography, or through an affinity column coupled with the transcriptional regulator of the invention. Besides being used to purify or detect the transcriptional regulators of the present invention, the antiobodies of the present invention can be applied to antibody therapy.

The present invention also relates to a screening method for a compound that binds to transcriptional regulators of the present invention. The screening method of the present invention includes steps of contacting a transcriptional regulator of the present invention with a test sample and selecting a compound that has binding activity for the transcriptional regulator of the present invention. Test samples used for the screening include, but

are not limited to, a cell extract, a supernatant of the cell culture, a library of synthetic low molecular weight compounds, a purified protein, an expression product of a gene library, and a library of synthetic peptides.

Methods well known to one skilled in the art for isolating a compound binding to a transcriptional regulator of the present invention using the regulator are as follows. A protein that binds to a transcriptional regulator of the present invention can be screened by West-western blotting comprising steps of generating a cDNA library from the cells expected to express the protein that binds to a transcriptional regulator of the present invention (e.g., testis tissue cell and tumor cell lines HL-60, HeLa S3, Raji, and SW480) using a phage vector (λgt11, ZAP, etc.), allowing the cDNA library to express on the LB-agarose plate, fixing the expressed proteins on a filter, reacting them with the transcriptional regulator of the present invention purified as a biotin-labeled protein or a fusion protein with GST protein, and detecting plaques expressing the protein bound to the regulator on the filter with streptavidin or anti-GST antibody (Skolnik, E. Y., Margolis, B., Mohammadi, M., Lowenstein, E., Fisher, R., Drepps, A., Ullrich, A. and Schlessinger, J. (1991), Cloning of PI3 kinase-associated p85 utilizing a novel method for expression/cloning of target proteins for receptor tyrosine kinases, Cell, 65:83-90). Alternatively, the method comprises expressing in yeast cells a transcriptional regulator of the present invention which is fused with SFR or GAL4 binding region, constructing a cDNA library in which proteins are expressed in a fusion protein with the transcription activation site of VP16 or GAL4 from the cells expected to express a protein that binds to the transcriptional regulator of the present invention,

introducing the cDNA library into the above-described yeast cells, isolating the cDNA derived from the library from the detected positive clones, and introducing and expressing it in *E. coli*. (If a protein that binds to the transcriptional regulator of the present invention is expressed, a reporter gene is activated by the binding of the two proteins. The positive clones can then be identified.) This method can be performed using Two-hybrid system (MATCHMAKER Two-Hybrid System, Mammalian MATCHMAKER Two-Hybrid Assay Kit, MATCHMAKER One-Hybrid System (all from Clontech); HybriZAP Two-Hybrid Vector System (Stratagene) or in accordance with Dalton, S. and Treisman R. (1992), Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element, Cell, 68:597-612). Another method is to apply a culture supernatant or a cell extract from the cells suspected to express a protein which binds to the transcriptional regulator of the present invention onto an affinity column coupled with the transcriptional regulator of the present invention, and purify the protein specifically bound to the column.

Also well known to one skilled in the art are a method of screening molecules that bind to a transcriptional regulator of the present invention by reacting the immobilized transcriptional regulator of present invention with a synthetic compound, natural substance bank, or a random phage peptide display library, and a method of screening low molecular weight compounds, proteins (or their genes), or peptides which bind to a transcriptional regulator of the present invention by utilizing the high-throughput technique of combinatorial chemistry (Wrighton, N. C., Farrell, F. X., Chang, R., Kashuyap, A. K., Barbone, F. P., Mulcahy, L. S., Johnson, D. L., Barrett, R. W., Jolliffe, L. K., Dower, W. J., Small peptides as potent

mimetics of the protein hormone erythropoietin, Science (UNITED STATES) Jul. 26, 1996, 273:458-464; Verdine, G. L., The combinatorial chemistry of nature, Nature (ENGLAND), Nov. 7, 1996, 384:11-13; Hogan, J. C. Jr., Directed combinatorial chemistry, Nature (ENGLAND), Nov. 7, 1996, 384:17-19). The compounds thus isolated, which bind to a transcriptional regulator of the present invention, may be used to treat cancer or other proliferative diseases. When the compounds isolated by the screening method of the present invention are used as pharmaceuticals, they can be formulated by a known pharmacological process. For example, they can be administered to a patient with pharmaceutically acceptable carriers and vehicles (e.g., physiological saline, vegetable oil, a dispersant, a surfactant, and a stabilizer). The compounds may be percutaneously, intranasally, transbronchially, intramuscularly, intravenously, or orally administered, depending on their properties.

The present invention also relates to DNA specifically hybridizing with DNA coding a protein of the present invention and having at least 15 nucleotides. As used herein, "specifically hybridizing" means that no cross-hybridization occurs between DNA encoding other proteins under conditions of moderate stringency. Such DNA may be used as a probe for detecting and isolating the DNA encoding the protein of the present invention, and as a primer for amplifying the DNA encoding the protein of the present invention.

An "isolated nucleic acid" is a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for

example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of different (i) DNA molecules, (ii) transfected cells, and (iii) cell clones: e.g., as these occur in a DNA library such as a cDNA or genomic DNA library.

The term "substantially pure" as used herein in reference to a given polypeptide means that the polypeptide is substantially free from other biological compounds, such as those in cellular material, viral material, or culture medium, with which the polypeptide was associated (e.g., in the course of production by recombinant DNA techniques or before purification from a natural biological source). The substantially pure polypeptide is at least 75% (e.g., at least 80, 85, 95, or 99%) pure by dry weight. Purity can be measured by any appropriate standard method, for example, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

A "conservative amino acid substitution" is one in which an amino acid residue is replaced with another residue having a chemically similar side chain. Families of amino acid residues having similar side chains have been defined

in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

As used herein, "percent identity" of two amino acid sequences or of two nucleic acids is determined using the algorithm of Karlin and Altschul (*Proc. Natl. Acad. Sci. USA* 87:2264-2268, 1990), modified as in Karlin and Altschul (*Proc. Natl. Acad. Sci. USA* 90:5873-5877, 1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al. (*J. Mol. Biol.* 215:403-410, 1990). BLAST nucleotide searches are performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches are performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a reference polypeptide. To obtain gapped alignments for comparison purposes, Gapped BLAST is utilized as described in Altschul et al. (*Nucleic Acids Res.* 25:3389-3402, 1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) are used. See http://www.ncbi.nlm.nih.gov.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. In case of conflict, the

present application, including definitions, will control. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference. The materials, methods, and examples are illustrative only and not intended to be limiting. Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

Brief Description of the Drawings

Figure 1 compares the domain of BAZ (BAZ1α) with those of other proteins. In Fig. 1A, the bromodomain of BAZ (BAZ1α) is compared with that of TFIID from yeast, CCG1 from human, p300, and CBP. In Fig.1B, C4HC3 Zn finger of BAZ (BAZ1α) is compared with those of U13646, retinoblastoma binding protein RBP2, two species of MOZ, p300, and CBP. The conserved amino acids, cysteine and histidine, are indicated by "*." In both Figs. 1A and B, identical amino acids are represented by reverse-contrast letters, and similar amino acids are represented by emphasized letters (also netted).

Figure 2 shows a chromosome map of BAZ (BAZ1α). Figure 2A shows assignments of chromosome 14 based on the analysis of a monochromosome hybrid cell panel using primers B (SEQ ID NO:6) and M (SEQ ID NO:7). The numbers 1 to Y in the figure refer to chromosome numbers, and TH refers to total human chromosomes. The 111 bp product was specifically amplified in the cell line GM10479, monochromosome1 for a human chromosome14. Figure 2B depicts the location of BAZ (BAZ1α) on chromosome 14 as determined by Genebridge 4 radiation hybrid panel analysis.

Figure 3 shows the expression of BAZ (BAZ1α) in normal tissues (Lane 1, heart; Lane 2, brain; Lane 3, placenta; Lane 4, lung; Lane 5, liver; Lane 6, skeletal

muscle; Lane 7, kidney; Lane 8, pancreas; Lane 9, spleen; Lane 10, thymus; Lane 11, prostate; Lane 12, testis; Lane 13, ovary; Lane 14, small intestine; Lane 15, colon (mucous lining); and Lane 16, leukocytes in the peripheral blood). The bottom of the figure shows control bands using actin probes.

Figure 4 shows the expression of BAZ (BAZ1α) in carcinoma. Figure 4A depicts the Northern blot analysis in the carcinoma cell lines (Lane 1, promyelocytic leukemia HL-60; Lane 2, HeLa S3 cells; Lane 3, chronic myelocytic leukemia K-562; Lane 4, lymphoblastic leukemia MOLT-4; Lane 5, Burkitt's lymphoma Raji; Lane 6, large intestine adenocarcinoma SW480; Lane 7, lung carcinoma A549; and Lane 8, melanoma G361). Figure 4B shows the RT-PCR analysis of primary lung carcinoma of Lane 10. The top panel shows the amplified product of 554 bp from BAZ (BAZ1α) gene using primers U and N, and the bottom shows the amplified product of 442 bp from G3PDH gene using primers G3U and G3L. In the figures, L refers to normal human lung; T, to normal human testis; and G, to normal human genomic DNA.

Figure 5 compares the conserved domains among BAZ (BAZ1α), BAZ2α, and U13646.

Figure 6 shows alignments of the domain of BAZ2α and that of other proteins. In Fig. 6A, the bromodomain of BAZ2α (BAZ2 in the figure) is aligned with BAZ (BAZ1α) human CCG1, PCAF, and CBP. In Fig. 6B, C4HC3 Zn finger of BAZ2α is aligned with BAZ (BAZ1α), U13646, retinoblastoma binding protein RBP2, 2 zinc fingers of MOZ, and p300. The conserved cysteine and histidine are indicated by "*."

Figure 7 shows a chromosome map of BAZ2α. Figure 7A shows assignments of chromosome 12 based on the analysis of a monochromosome hybrid cell panel using primers D (SEQ ID NO:16) and E (SEQ ID NO:17). The numbers 1 to Y in the

figure refer to chromosome numbers. The product of 132 bp was specifically amplified in the cell line GM10868a, a monochromosome for human chromosome 12. Figure 7B depicts the location of BAZ2 (BAZ2α) on chromosome 12 as determined by Genebridge 4 radiation hybrid panel analysis.

Figure 8 shows the expression of BAZ2α in normal tissues (A: Lane 1, heart; Lane 2, brain; Lane 3, placenta; Lane 4, lung; Lane 5, liver; Lane 6, skeletal muscle; Lane 7, kidney; Lane 8, pancreas; Lane 9, spleen; Lane 10, thymus; Lane 11, prostate; Lane 12, testis; Lane 13, uterus; Lane 14, small intestine; Lane 15, colon (mucous lining); and Lane 16, leukocytes in the peripheral blood). Figure 8B shows controls using actin probe.

Figure 9 compares the conserved domains of BAZ2β, BAZ1α, BAZ2α, U13646, and BAZ1βS. Each has at least five domains. BAZ1α lacks domain I. The figures on the bars of BAZ2β, BAZ1α, U13646, and BAZ1βS represent the percentage of the homology with BAZ2α. The values in the LH domain represent the percentage of the homology with leucine residues. Black bands in the LH domain indicate where the LXXLL motif is present in all three BAZ genes. LH, ZF, and BD represent leucine-rich helix domain, C4HC3 zinc finger, and bromodomain, respectively.

Figure 10 compares the amino acid sequence of LH domain in BAZ2β with those of corresponding domains in other proteins. The positions of conserved leucine residues are indicated by arrows on the sequences. LXXLL motifs are boxed.

Figure 11 shows a chromosome map of BAZ2β. The product of 147 bp was specifically amplified in the cell line as a monochromosome for human chromosome 2. This product was amplified by PCR using primers nb7n and nb7ee. The numbers 1 to Y in the figure indicate chromosome

numbers. The location of BAZ2β on chromosome 2 was determined by Genebridge 4 radiation hybrid panel analysis.

Figure 12 shows an analysis of the expression of BAZ2β in normal tissues, carcinoma cell lines, and fetal tissues (Lane 1, heart; Lane 2, brain; Lane 3, placenta; Lane 4, lung; Lane 5, liver; Lane 6, skeletal muscle; Lane 7, kidney; Lane 8, pancreas; Lane 9, spleen; Lane 10, thymus; Lane 11, prostate; Lane 12, testis; Lane 13, uterus; Lane 14, small intestine; Lane 15, colon (mucous lining); Lane 16, leukocytes in the peripheral blood; Lane 17, fetal brain; Lane 18, fetal lung; Lane 19, fetal liver; Lane 20, fetal kidney; Lane 21, acute leukemia HL-60; Lane 22, HeLa S3 cells; Lane 23, chronic myelocytic leukemia K-562; Lane 24, lymphoblastic leukemia MOLT-4; Lane 25, Burkitt's lymphoma Raji; Lane 26, large intestine adenocarcinoma SW480; Lane 27, lung carcinoma A549; and Lane 28, melanoma G361). The sizes of the transcripts are indicated on the right side of the figure.

Figure 13 shows the alignments of variable portions of BAZ1βS and BAZ1βL.

Figure 14 shows the alignments of N terminal amino acid sequences from BAZ1βS and three other members of the BAZ family. The residues with 50% or more sequence homology are indicated by dark shadowed boxes, and those with 50% or more sequence similarity, by light shadowed boxes. Conserved LXXLL motifs and C4HC3 zinc fingers are indicated on the alignments. Conserved leucine residues in the surrounding region of the LXXLL motif are indicated. The location of a bromodomain motif is indicated by a black line on the alignments.

Figure 15 shows the alignments of the amino acid sequences from BAZ1βS and three other members of the BAZ family (continued from Fig. 14).

Figure 16 shows the alignments of the amino acid sequences from BAZ1βS and three other members of the BAZ family(continued from Fig. 15.

Figure 17 shows the alignments of the amino acid sequences from BAZ1βS and three other members of the BAZ family (continued from Fig. 16).

Figure 18 shows the alignments of the amino acid sequences from BAZ1βS and three other members of the BAZ family (continued from Fig. 17).

Figure 19 shows a chromosome map of BAZ1β. Figure 19A shows mapping of BAZ1β on chromosome seven by monochromosome hybrid cell line panel analysis. A product of 156 bp was observed to be amplified in the cell line GM10791 by PCR using primers nb3S and nb3T. The numbers 1 to Y in the figure indicate chromosome numbers. Figure 19B shows the location of BAZ1β on chromosome seven as determined by Genebridge 4 radiation hybrid panel analysis. BAZ1β is located between 7q11-21 markers D7S489 and D7S669.

Figure 20 shows the expression analysis of BAZ1β in normal tissues. In Fig. 20A, the BAZ1β probe is hybridized with two transcripts in a wide range of tissues (Lane 1, heart; Lane 2, brain; Lane 3, placenta; Lane 4, lung; Lane 5, liver; Lane 6, skeletal muscle; Lane 7, kidney; Lane 8, pancreas; Lane 9, spleen; Lane 10, thymus; Lane 11, prostate; Lane 12, testis; Lane 13, uterus; Lane 14, small intestine; Lane 15, colon (mucous lining); and Lane 16, leukocytes in the peripheral blood). Figure 20B shows controls using an actin probe. In Fig. 20B, the blot in Fig. 20A was used to rehybridize with the actin probe.

Figure 21 shows the regions within BAZ2β which are covered by expression clones. Conserved domains (shadowed boxes) and LXXLL motifs (black lines) are indicated. Positions of the first and the last amino acids of each

domain are indicated on the bar. Clone 1 covers amino acids 1-190; clone 9, amino acids 1241-1584; and clone 11, amino acids 1500-1970.

Figure 22 is a photograph of electrophoretic patterns showing SDS-PAGE analysis of GST protein (Lane 1, cell lysate (BAZ2β.1); Lane 2, flow through fraction (BAZ2β.1); Lane 3, purified fusion protein (BAZ2β.1); Lane 4, cell lysate (BAZ2β.11); Lane 5, cell lysate (BAZ2β.9); Lane 6, flow through fraction (BAZ2β.11); Lane 7, flow through fraction (BAZ2β.9); Lane 8, purified protein (BAZ2β.11); and Lane 9, purified protein (BAZ2β.9). The positions of molecular weight markers are indicated on the right (kDa).

Figure 23 is a photograph of electrophoresis showing Western analysis of purified GST-fusion protein (Lane 1, GST; Lane 2, GST-BAZ2β.1; Lane 3, GST-BAZ2β.9; and Lane 4, GST-BAZ2β.11).

Best Mode for Implementing the Invention

The present invention is further illustrated with reference to the following examples, but is not to be construed to be limited thereto.

Example 1 Isolation and analysis of BAZ gene

(1) Identification of novel genes each containing a bromodomain

The EST database was searched by means of BLAST using the DNA sequence that encodes the 5′ bromodomain motif of the RING3 gene (SEQ ID NO:3) (Beck, S. et al., (1992), DNA Sequence, 2:203-210), and a number of ESTs identical to the probe sequence were retrieved. The following experiment was then performed for one of those EST, H70181. H70181 has the highest homology to transcription activator GCN5 of

yeast (Georgakopoulos, T. and Thireos, G. (1992), EMBO J., 11(11):4145-4152) or human (Candau, R. et al., (1996), Mol. Cell. Biol., 16(2):593-602).

(2) Isolation of a full-length sequence

To clone a full-length sequence of EST H70181, PCR primers were designed; primer U, SEQ ID NO:4/ AGAAAAAGACAATCTCCAGAGCA, and primer L, SEQ ID NO:5/ GCTGTCATCATGTCGTACCAATTC. The specific product of 129 bp obtained from testis cDNA was amplified by RT-PCR using said primers. The amplified product was directly purified through a QIA Quick (Qiagen) purification column. This PCR product was used as a probe for screening the testis cDNA library (Clontech; HL3023a). The probe was labeled with [α-^{32}P]dCTP by random priming, and purified using a Chromaspin 10 Column (Clontech); the cDNA clone obtained was used to re-screen the library. This process was repeated until a series of overlapped clones was obtained, and thus a full-length sequence was obtained. The isolated sequence was 5,934 bp in total. The isolated gene was designated "BAZ" (Bromodomain, Atypical Zinc finger). BAZ has an open reading frame (ORF) (SEQ ID NO:1) coding for 1,674 amino acids from the nucleotide positions 125-5147. The ORF is followed by 787bp of a 3' untranslated region and terminated with a poly-A tail. The polyadenylation signal (AATAAA) is located at 21 bp upstream from the poly-A tail. The nucleotide sequence together with the deduceded amino acid sequence therefrom is shown in SEQ ID NO:2.

The filter screening of the library was performed in ExpressHyb hybridization solution (Clontech) at 65°C for 1 hour. The filter was then washed until a final stringency of 1 x SSC and 0.1% SDS at 65°C was attained. All the sequencing was performed on automated sequencing apparatus

ABI 377 (Perkin Elmer, Norwalk, CT), utilizing ABI dye terminator chemistry.

(3) Identification of the homology and motifs characterizing the transcription factors

A protein database search using the amino acid sequence of BAZ revealed that a protein encoded by a continuous 2.2 Mb gene sequence of the chromosome III of C. Eelegans (Wilson, R. et al., (1994), Nature, 368:32-38) is most similar (46% similarity and 23% identity). The same regions having similarity were found in various transcription factors such as the 250 KD subunit of TFIID (Ruppert, S., Wang, E., and Tjian, R. (1993), Nature 362:175-179) and p300/CBP (Eckner, R. et al., (1994), Genes Dev., 8(8):869-884; Chrivia, J. C. et al., (1993), Nature, 365:855-859). A motif search of amino acid sequences in the PROSITE database on DNASIS (HITACHI Software Engineering Co.) identified a single bromodomain (amino acid residues 1569-1627 of SEQ ID NO:1). The sequence of this bromodomain, together with those of other bromodomains, is shown in Fig. 1A. A BLAST search using C4HC3 Zn finger (C4HC3ZF), which is the motif conserved among a great variety of proteins such as U13646, identified retinoblastoma binding protein RBP2 (Fattaey, A. R. et al., (1993), Oncogene, 8:3149-3156), MOZ (Borrow, J. et al., (1996), Nature Genet., 14:33-41), and p300/CBP(Koken, M. H. et al., (1995), CR Acad. Sci. III, 318:733-739), a motif of 45 amino acids (corresponding to amino acid residues 1269-1313 of SEQ ID NO:1). C4HC3ZFs present in these genes are shown in Fig. 1B. The function of BAZ as a transcriptional regulator is implied by the fact that it is similar to several transcriptional regulators, in particular, a bromodomain motif conserved together with C4HC3ZF and

p300/CBP. The similarity of BAZ to p300/CBP is not limited to C4HC3 zinc finger and bromodomain regions; well conserved regions are also found adjacent to the bromodomain. Homology was not found between BAZ and histone-acetyltransferase domain, and between BAZ and other domains in which p300/CBP is present. However, BAZ potentially has HAT activity since the histone-acetyltransferase domain is not well conserved among proteins.

Several sorts of sequence motifs characterized by the nuclear proteins were identified at 11 sites by employing the PSORT program (http://psort.nibb.ac.jp) utilizing a wide variety of conserved nuclear localization sequences.

(4) Mapping of BAZ

Primers B SEQ ID NO:6/ AACACAAGTGAAGCAAAAGCTGGA and M SEQ ID NO:7/ GTGGTGTGCTAACTTGGTCCACAT (obtained from the 3′ end of the gene) were used to amplify DNA obtained from each of the 24 monochromosomes of human/rodent somatic cell lines available from Coriell Cell Respositories, New Jersey (Dubois, B. L. and Naylor, S. (1993), Genomics, 16:315-319). The expected product of 111 bp was amplified only from GM10479, a monochromosomal cell line for human chromosome 14 (see Fig. 2A). Primers B (SEQ ID NO:6) and M (SEQ ID NO:7) were subsequently used for PCR onto a GeneBridge 4 radiation hybrid panel (Research Genetics, Huntsville, Alabama). The binary codes generated by assessing whether each hybrid is positive or negative for amplification were compared with the analogous codes for the markers constituting a framework map, using the server located at http://www-genome.wi.mit.edu./cgi-bin/contig/rhmapper.pl. This step was repeated using primers W: SEQ ID NO:8/ CCCATCGTGAGTCAAGAGTGTCTGT and X: SEQ ID NO:9/

CTCGCTTCTACCTTTTTATTGGCT (from the 5' end of the gene). Based on the pattern obtained from this panel by identifying the amplification in the panel, BAZ was proved to be located on the 14q between the two markers D14S730 and D14S75 (see Fig. 2B).

(5) Analysis of BAZ expression in the normal tissues

The probe as synthesized in Example 1(2) by amplifying cDNAs from testis with PCR using the above-described primers U (SEQ ID NO:4) and L (SEQ ID NO:5) was used for Northern blot analysis of 16 panels of normal tissues. The probe hybridized with a single species of mRNA of 7.0 kb, which corresponds to the length of the ORF identified from the sequence of the gene. Though this transcript was expressed in almost all the tissues at very low levels, it was expressed in testis at a relatively high level (see Fig. 3). The transcript was not detectable in brain, lung, liver, kidney, and colon (it was possibly expressed at non-detectable levels). A slightly smaller transcript of 6.5 kb was also detected only in testis at a low level. Since the cells divide more vigorously in the testis than any other tissue examined, the expression pattern is thought to correspond to the role for BAZ during active proliferation. Hybridization for Northern blot analysis was performed in Express Hyb hybridization solution (Clontech) at 65°C for 1 hour. The filters were washed until the final stringency reached 1 x SSC and 0.1% SDS at 65°C. Imaging was performed using a Fuji BAS Image Analyzer (Fuji Photo Film).

(6) Analysis of BAZ expression in tumor

That BAZ is highly expressed in the testis suggests the possibility of its high level expression in vigorously

proliferating tumors. Thus, Northern analysis of eight panels of tumor cell lines was carried out, using the same probe as used in Example 1(5). As a result, the transcript of 7.0 kb alone was hybridized with the probe as when the normal tissue was used. Compared with most normal tissues, however, the transcriptional levels are remarkably higher in most of the tumor cell lines (see Fig. 4A). Specifically, the expression was higher in the tumor cell lines HL-60, HeLa S3, Raji, and SW480. In contrast, the expression levels in K-562, MOLT-4, A549, and G361 are almost the same as those in normal tissues.

RT-PCR was used to examine the expression of BAZ in the primary lung carcinomas as shown in Table 1.

Table 1

Sample No.	Patients age	sex	Carcinoma
1	49	male	papillary adenocarcinoma
2	63	male	papillary adenocarcinoma (moderately differentiated)
3	60	male	papillary adenocarcinoma (poorly differentiated)
4	70	male	squamous cell carcinoma (fusiform cell variant)
5	76	male	papillary adenocarcinoma
6	65	male	large lung cell carcinoma (moderately differntiated)
7	77	male	squamous cell carcinoma (poorly differentiated)
8	45	male	acinic adenocarcinoma
9	50	male	carcinoid tumor
10	66	male	choriocarcinoma

Each of the 10 samples was amplified using the primers G3U SEQ ID NO:10/TCATCATCTCTGCCCCCTCTGTCTG and G3L SEQ ID NO:11/GACGCCTGCTTCACCACCTTCTTG, which are the primers for amplifying 442 bp of a house-keeping gene G3PDH, and the primers U, SEQ ID NO:4 and N, SEQ ID NO:12/ TCATGTGGTCAATCAATTGTTTGT, which are primers for BAZ (see Fig. 4). G3PDH was used to determine that an equal amount of mRNA was present in each sample.

The primers for BAZ were selected to specifically amplify the cDNA but not genomic DNA. The amplified product was definitely detected in the sample from the testis and

the two lung tumors, but not from the other eight samples from the lung tumor or the normal lungs.

RT-PCR was performed according to a standard technique in which total RNA was extracted according to the AGPC method (Chomczynski, P. and Sacchi, N. (1987), Analytical Biochem., 162:156-159), then single-stranded DNA was synthesised with an oligo (dT15) primer and MMLV reverse transcriptase, a part of which was used for the RT-PCR. The RT-PCR was performed using AmpliTaq gold (ABI), with 27 cycles of annealing at 60°C to amplify G3PDH and 33 cycles of annealing at 55°C to amplify BAZ. The conditions for hybridization and imaging were the same as in Example 1(5).

Example 2 Isolation and analysis of BAZ2α gene

(1) Identification of a novel gene containing a bromodomain and isolation of its full-length sequence

The DNA encoding the bromodomain of BAZ is highly homologous to that of GCN5. The DNA sequence encoding the bromodomain motif of human GCN5 gene (Candau, R. et al., (1996), Mol. Cell. Biol., 16 Q):593-602) was used to search the EST database using BLAST. The Motif search was performed using PROSITE. Proteins were compared using Bestfit in GCG. The nuclear transport signal was identified using PSORT. As a result, a number of ESTs were found to be identical to the probe sequence. Among them, an EST (Accession Number: N76552) obtained from a fetal liver/spleen cDNA library proved to be a novel gene.

To start cloning the full-length sequence of EST N76552, PCR primers were designed to amplify a particular product of 91 bp from the testis cDNA library; primer NB16U (SEQ ID NO:15/ TGACTCTGAAGTAGGCAAGGCTGG) and primer NB16L (SEQ ID NO:16/ CTTGCCTCACAGATTGGCCTGT). The PCR product was used as a probe to screen the testis cDNA library (Clontech;

HL3023a). The amplified product was directly purified through a QIA Quick (Qiagen) purification column. The cDNA clone having sequences corresponding to EST was used to re-screen the library.

This process was repeated until a series of overlapped clones having a full-length sequence of the complete coding region was obtained. All the sequencing was performed with automated sequencing apparatus ABI 377 (Perkin Elmer, Norwalk, CT), utilizing ABI dye terminator chemistry. As a result, a continuous sequence consisting of 9,408 bp nucleotides in total size was generated. Theoretical translation of this sequence showed a presence of methionine codon at the nucleotide position of 740. An open reading frame (ORF) coding 1878 amino acids starts from this position and terminates at the nucleotide position 6373. The ORF is followed by a 3' untranslated region consisting of at least a 3 kb nucleotide sequence. The nucleotide sequence of the cDNA obtained is shown in SEQ ID NO:14, and the amino acid sequence deduced from the cDNA is shown in SEQ ID NO:13. The isolated clone was designated BAZ2α.

(2) Identification of the homology and motifs characteristic of transcription factors

Like BAZ, BAZ2α was shown to have the highest homology with the protein encoded by a C. elegans bromodomain gene U13646, the gene forming a part of a continuous 2.2 Mb segment of chromosome III of C. elegans (Wilson, R. et al., (1994), Nature, 368:32-38) by searching the protein databases with the amino acid sequence of BAZ2α. The regions which showed similarity were identified using various transcription factors such as the 250 KD subunit of TFIID (Ruppert, S., Wang, E. and Tjian, R. (1993), Nature

362:175-179) and p300/CBP (Eckner, R. et al., (1994), Genes Dev., 8(8):869-884; Chrivia, J. C. et al., (1993), Nature, 365:855-859). The bromodomain was located between the amino acid residues 1788 and 1846. The alignments of BAZ2α, BAZ, and U13646 are shown in Fig.5. The alignments of the sequence of the BAZ2α bromodomain and those of other bromodomains are shown in Fig. 6A. Moreover, a single motif consisting of 45 amino acids (amino acid residues 1652-1696) was identified. This motif codes C4HC3 Zinc finger (C4HC3ZF), a motif conserved among a large number of proteins such as BAZ, U13646, retinoblastoma binding protein RBP2 (Fattaey, A. R. et al., (1993), Oncogene, 8:3149-3156), MOZ (Borrow, J. et al., (1996) Nature Genet., 14:33-41), and p300/CBP (Koken, M. H. et al., (1995) CR, 4 cad. Sci. III, 318:733-739) by BLAST searching. The alignments of C4HC3ZF from these genes are shown in Fig. 6B. BAZ2α resembles BAZ, which suggests the possibility that the two proteins closely relate and form a part of a protein family having a similar function. Like BAZ, BAZ2α wholly resembles several transcription factors and has C4HC3ZF and bromodomain motifs conserved among p300/CBP and TIF1, especially indicating that BAZ is likely to function as a transcriptional regulator.

An LXXLL motif, which is believed to be required for mediating transcription induced by nuclear receptors (Torchia, J. et al., (1997), Nature, 387:677-684; Herry, D. M. et al., (1997), Nature 387:733-736), is located at amino acid residue 872. PROSITE motif searching revealed that this motif was located at the 3′ end of the leucine zipper (amino acid residues 852-873). The relative locations of LXXLL, C4HC3, and bromodomain motifs in BAZ2α are remarkably similar to those of U13646 and BAZ (Fig. 5). Furthermore, in either case, the LXXLL motif is located behind the helix

structure characterized by conserved lysine residues existing at regular intervals.

(3) Mapping of BAZ2α

To locate BAZ2α on the chromosome, PCR primers D (SEQ ID NO:17/ TTGCCGTATTTGGCTGGTATC) and E (SEQ ID NO:18/ CATAGAGAAGAGGGCAGGGTTGA), which amplify a fragment of 132 bp, were used to amplify the DNA from each of the 24 monochromosomes of human/rodent somatic cell lines (Dubois, B. L. and Naylor, S. (1993), Genomics, 16:315-319) obtained from Coriell Cell Respositories (New Jersey). The BAZ2α-containing region was identified using 91 GeneBridge 4 radiation hybrid panels (Walter M. A. et al., (1994), Nature Genetics, 7:22-28). These panels were screened by PCR using primers D and E again. The binary codes generated by assessing whether each hybrid is positive or negative for amplification were compared with the analogous codes for the markers constituting a framework map, using the server located at http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl. BAZ2α was thus proved to be located 12q24.3-ter from D12S367(see Fig. 7B).

(4) Analysis of the BAZ2α expression

The probe (a 481 bp fragment of BAZ2α gene) prepared by amplifying the DNA from one of the clones obtained from the testis cDNA library (Clontech)in Example 2(1) using primers gt10F (SEQ ID NO:19/ CTTTTGAGCAAGTTCAGCCT) and NB16N (SEQ ID NO:20/ GTCGGCTTCTTCATTTCCTCCA) was used for Northern analysis of 16 panels of normal tissues (Clontech). The probe was labeled with [α-^{32}P]dCTP by random priming and purified using a Chromaspin 10 column (Clontech). Hybridization for Northern analysis and library filter screening were performed in the ExpressHyb hybridization

solution (Clontech) at 65°C for 1 hour. The filters were then washed until the final stringency reached 0.5 x SSC and 0.1 % SDS. Imaging was performed using a Fuji BAS Image Analyzer. The result showed that this probe was hybridized with a single species of mRNA of 10.5 kb in almost all the tissues; this length corresponds to that of ORF identified from the gene sequence (Fig. 8). The transcript was expressed in almost all the tissues at a low level. Another transcript of 9.0 kb was detected and was primarily expressed in the testis. This band survived the high-stringent wash. The second transcript is thought to be an alternatively spliced form of, or a different gene closely related to, BAZ2α.

Example 3 Isolation and analysis of BAZ2β gene

(1) Identification of a novel gene containing a bromodomain and isolation of its full-length sequence

A BLAST search was performed against the EST databases using the various nucleotide sequences containing a known bromodomain motif. Several ESTs which may encode the bromodomain gene were identified based on the result of the search using the nucleotide sequence of the SMAP gene (Nielsen, M. S. et al., (1996), Biochem. Biophys. Acta). Among them, an EST (Gnbank Accession Number: AA015589) obtained from a retinal cDNA library was proved to be a novel gene, the protein deduced from which has the highest homology with BAZ2α.

Its full-length nucleotide sequence was isolated. The full-length gene for EST AA015589 was cloned as follows. First, PCR primers NB7U (SEQ ID NO:23/ CTGACTGAAATGGAAACTCATGAGG) and NB7L (SEQ ID NO:24/ CTAGAGCAAAGGTTTCAAGGTTTGG) were designed to obtain the specific product of 165 bp from the testis cDNA. The

amplified product was directly purified with a QIA Quick (Qiagen) purification column. The PCR product was used as a probe to screen the testis cDNA library, and the cDNA clone containing the EST sequence was used to re-screen the library. This process was repeated until the nucleotide sequence covering the whole coding region of the gene was obtained by assembling the clones. As a result, a nucleotide sequence consisting of 7,585 bp in total was obtained. This full-length sequence contains an open reading frame (ORF) consisting of 1972 amino acids (6,282 nucleotides) with ATG at the nucleotide position 367 as the initiation codon, followed by 3' UTR of 1303 bp. SEQ ID NO:22 shows the nucleotide sequence of the cDNA thus obtained, and SEQ ID NO:21 shows the amino acid sequence deduced from the nucleotide sequence. All the sequencing was performed on automated sequencing apparatus ABI 377 (Perkin Elmer, Norwalk, CT), utilizing ABI dye terminator chemistry. Hybridization for the library filter screening was performed using ExpressHyb hybridization solution (Clontech) at 65°C for 1 hour. The filters were washed until the final stringency reached 0.5 x SSC and 0.1% SDS. Subsequently, the filters were subjected to autoradiography at -70°C for 1 to 3 days to intensify their signals.

(2) Homology and the characteristics of the motifs of the transcriptional regulator

The motifs of the protein encoded by the gene obtained were searched for in PROSITE. The proteins were compared using Bestfit from GCG. A nuclear transport signal was identified through PSORT (http://psort.nibb.ac.jp/form.html).

As for BAZ and BAZ2α, a database search based on the amino acid sequences predicted from the registered genes

showed that this gene has the greatest similarity to the protein encoded by the bromodomain gene U13646 from the nematode (C. elegans). This nematode bromodomaingene corresponds to a portion of a 2.2 Mb segment derived from chromosome III of the nematode (C. elegans) (Wilson, R. et al., (1994), Nature, 368:32-38). The gene, however, shows homology to BAZ and BAZ2α to a larger extent. Actually, the similarity of the protein encoded by this gene to BAZ and BAZ2α suggests the possibility that these three proteins are closely related to one another, and, moreover, that they are a part of a broader family of proteins with similar functions. This gene was designated BAZ2β (for bromodomain, atypical zinc finger), since it has the greatest association with BAZ2α. BAZ was also renamed BAZ1α. The amino acid sequence of BAZ2β is shown in Fig. 9 together with those of BAZ1α, BAZ2α, U13646, and BAZ1βS described below. At least five regions or domains can be identified from the sequences. The first domain (I) is not present in BAZ1α, but is in the other three proteins. The existence of a leucine-rich helical structure (LH) was predicted from the analysis of the next domain. LXXLL motif is present at the central part of this domain on all BAZs except U13646. This motif potentially confers the interaction with the nuclear receptors on the protein (Torchia, J. et al., (1997), Nature, 387:677-684; Heery, D. M. et al., (1997), Nature, 387:733-736). Both domains II and III are highly conserved, suggesting their functional importance. Each protein has a highly conserved C4HC3 zinc finger (Aasland, R. et al., (1995), Trends Biochem. Sci., 20:56-59; Koken, M.H. et al., (1995), CR Acad. Sci. III, 318:733-739; Saha, V. et al.,, (1995), Proc. Natl. Acad. Sci., 92:9737-9741) and a bromodomain. In addition, a conserved region is found upstream from the zinc-finger motif, and the region can also

be functionally important. Similarly, there are conserved sequences upstream from the bromodomain motif. Such conserved domains are aligned in Fig. 10. Like BAZ1α and BAZ2α, BAZ2β exhibits great similarity to several transcription factors and is thus expected to function as a transcription factor. Consistent with this function, estimation of the protein localization in the cell using the PSORT program revealed that BAZ2β has 19 consensus nuclear localization sequences (Robbins, J. et al., (1991), Cell, 64:615-23) in total.

(3) Chromosomal mapping of BAZ2β

To create a chromosome map of BAZ2β, primers nb7n (SEQ ID NO:25/ TGTTGCTGCATCACTTGTGTAGTT) and NB7ee (SEQ ID NO:26/ GGCATGACAATAATGTCTGCAAA) were prepared and used toamplify the DNA obtained from each of the 24 human/rodent monochromosomal somatic cell lines (Dubois, B. L. and Naylor, S. (1993), Genomics, 16:315-319). The amplification of the 147 bp fragment as expected PCR product indicated that the gene was likely to be located on human monochromosome 2 (Fig. 11). The locus region of BAZ2β was determined by use of 91 radiation hybrid panels of GeneBridge 4 (Walter, M.A. et al., (1994), Nature Genetics, 7:22-28). The hybrid panels were screened by PCR using primers nb7n and nb7ee again. The binary codes generated by assessing whether each hybrid is positive or negative for the amplification were compared with the analogous codes for the markers constituting a framework map, using the server located at http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl to identify the chromosomal locus of this gene. As a result, BAZ2β was confirmed to be located on chromosome 2q23-24 and between markers D2S1986 and G09369 (Fig. 11).

(4) Analysis of the BAZ2β expression

The cDNA probe containing the sequence correponding to nucleotide residues 1700-4000 was used for Northern analysis of 16 normal tissues, eight tumor cell lines, and four fetal tissues (Fig 12). The probe was labeled with [α-^{32}P]dCTP by random priming and purified on a Chromaspin 10 column (Clontech). Hybridization for Northern analysis was performed at 65°C for 1 hour using ExpressHyb hybridization solution (Clontech). Subsequently, the filters were washed at 65°C until the final stringency reached 0.5 x SSC and 0.1% SDS. Autoradiography was then performed at -70°C for 1 to 3 days to intensify the signals of the filters. This probe detected an mRNA of about 9.5 kb, a transcript whose size agreed with that of the ORF identified from the nucleotide sequence, in almost all the tissues examined. Besides this band, a transcript of about 6.5 kb was predominantly expressed in the testis. Since this band remained unchanged even after the high stringent wash (0.1 x SSC, at 65°C), it was considered to be specifically expressed. The second transcript could be an alternatively spliced product of BAZ2β, but no clone implying this event was found. It was also likely that expression of another gene closely related to BAZ2β was detected. In addition to these transcripts, several mRNAs were detected in most tissues. Such transcripts were considered to be derived from other genes each having an analogous sequence. Another analysis using another probe containing a bromodomain revealed the expression of the transcript of 6.5 kb only in the testis and of a 8.5 kb transcript in a wide range of tissues.

Example 4 Isolation and analysis of BAZ1β (BAZ1βS and BAZ1βL) genes

(1) Identification of novel genes containing a bromodomain and isolation of their full-length nucleotide sequences

A BLAST search was performed against the EST database using the nucleotide sequence of the bromodomain motif from human GCN5 gene (Candau et al., (1996), Mol. Cell. Biol., 16:593-602). Several ESTs possibly coding a number of bromodomain genes were identified. Among them, an EST (Gnbank accession Number: AA01307) derived from a retinal cDNA library was found to be a novel gene.

Its full-length sequence was isolated. The full-length gene for EST AA01307 was cloned as follows. First, PCR primers nb3U (SEQ ID NO:31/ TGGATGATGCTGAGGTGGATGA) and nb3L (SEQ ID NO:24/ GGGGTGCTGGATGACATCATAG) were designed to obtain a product of 184 bp specific to the primers from a testis cDNA library. The amplified product was directly purified using a QIA Quick (Qiagen) purification column. The PCR product was used as a probe to screen the testis cDNA library (Clontech HL3024a), and the cDNA clone containing the EST sequence was used to re-screen the library. This process was repeated after joining the clones. As a result, two types of nucleotide sequences were obtained and designated BAZ1β. The two sequences were further designated BAZ1βS for the shorter sequence and BAZ1βL for the longer one. The shorter sequence consisted of 5,561 nucleotides and encoded a protein of 1527 amino acids; the longer sequence consisted of 5,573 nucleotides and encoded a protein of 1531 amino acids, containing a tandem repeat of TACAGACCCTCC in one frame. This repeat gave rise to an insertion of four amino acids LLQT at position 658, which interestingly resulted in an additional LXXLL motif. BAZ1βS had four LXXLL motifs initiated at

positions 655, 658, 1000, and 1436, while BAZ1βL had five LXXLL motifs initiated at positions 655, 658, 663, 1004, and 1440. Figure 13 shows an alignment of the portions having multiple LXXLL motifs of BAZ1βS and BAZ1βL.

To determine whether the variability of the LXXLL motif is attributed to alteration of splicing or polymorphism, a pair of primers consisting of NB3KK (SEQ ID NO:33/ GAGTGCAGATAAGGGTGGCTTTTT) and NB3LL (SEQ ID NO:34/ CCAATTCACCATAGTCTTCGGCTA), which correspond to both sides of the variable region, was prepared and used to amplify genomic DNA and cDNA. As a result, these primers amplified a product of the same size from both of the templates. This implies the sequence variant is generated within an intron. Therefore, the variation of the sequence is probably caused by polymorphism. This may affect the interaction with the nuclear receptors. The nucleotide sequence of BAZ1βS cDNA thus obtained is shown in SEQ ID NO:28, and the deduced amino acid sequence of the protein encoded by the cDNA is shown in SEQ ID NO:27. The nucleotide sequence of BAZ1βL cDNA is also shown in SEQ ID NO:30, and the deduced amino acid sequence of the protein encoded by the cDNA is shown in SEQ ID NO:29. All the nucleotide sequences were determined with automated sequencing apparatus ABI 377, using ABI dye terminator chemistry. Hybridization for the filter screening of the library was performed in ExpressHyb hybridization solution (Clontech) at 65°C for 1 hour. The filters were washed at 65°C until the final stringency reached 0.5 x SSC and 0.1% SDS. Subsequently, the filters were autoradiographed at -70°C for 4 days to intensify the signals or autoradiographed for 4 hours with the Fuji BAS system.

(2) Homology and characteristics of the motifs of the transcriptional regulator

The motifs of the proteins encoded by BAZ1βS and BAZ1βL genes were searched in PROSITE. The proteins were compared using a MAP program of a BCM search launcher (http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html) under the default setting conditions; the output results were edited using a box shade program (http://ulrec3.unil.ch/software/BOX_form.htmlat). A nuclear transport signal was identified through PSORT (http://psort.nibb.ac.jp/form.html).

Several motifs characteristic of transcriptional regulators were found in both BAZ1βS and BAZ1βL. They were bromodomain, C4HC3 zinc finger (C4HC3ZF), and LXXLL motifs. LLXXLL motifs were present in the leucine-rich domain conserved among other BAZ family member protein genes and U13646 (Fig. 9). Although the importance of this domain has not been clarified, it can form a leucine zipper responsible for forming a dimer of the protein. It has been reported that such motifs are commonly found in the transcriptional regulators of eukaryotes (Busch and Sassone-Corsi, 1990) and that LXXLL motifs also interact with the nuclear receptors (Torchia et al., (1997), Nature, 387:677-684; Heery et al., (1997), Nature 387:733-736). That the predicted amino acid sequences have extensive similarity to several kinds of transcription regulators indicates the possibility that their genes function as transcriptional regulators. This is further supported by the fact that 13 nuclear localized consensus sequences (Robbins et al., (1991), Cell, 64:615-23) were found in total based on the prediction of the cellular localization of the proteins using the PSORT program. The predicted amino acid sequences exhibited the highest similarity to BAZ1α. They also showed similarity to

the proteins encoded by BAZ2α, BAZ2β, and C. elegans bromodomain gene U13646. Among the six domains, the first domain existed in BAZ2α, BAZ2β, and U13646, but not in BAZ1βS, BAZ1βL, or BAZ1α. Comparing the whole structures of these gene products, the region between domains II and III is the most similar to that of BAZ1α (Figures 14-18). Like other members of BAZ family, these gene products also have motifs that are present in the protein assumed to be encoded by nematode (C. elegans) bromodomain gene U13646 (Wilson et al., (1994) Nature, 368:32-38) that is identified by analyzing genome sequences of the genes. Alignment of the sequences of BAZ1βS, other members of the BAZ family, and U13646 reveals that the most highly conserved regions are located between the center and the C terminus of the sequences (Figs. 14-18). For U13646, this region is not depicted in the figures, and only N terminal region is aligned with that of BAZ1βS and BAZ1α.

(3) Chromosomal mapping of BAZ1β

To create a chromosome map of BAZ1β, primers nb3S (SEQ ID NO:35/ GAAACGGGAGGAGCTGAAAAAG) and nb3T (SEQ ID NO:36/ CCTTCAGGGGTATCCACCAATC) were prepared and used to amplify the DNA obtained from each of the 24 human/rodent monochromosomal somatic cell lines (Dubois, B. L. and Naylor, S. (1993), Genomics, 16:315-319). The expected PCR product of 156 bp was amplified from GM10791 from two distinct cell lines, suggesting that the BAZ1β gene is likely to be located on human chromosome 7 (Fig. 19A). The locus of BAZ1β was determined using 91 radiation hybrid panels of GeneBridge 4 (Walter, M. A. et al., (1994), Nature Genetics, 7:22-28). The hybrid panels were screened by performing PCR with primers nb3S and nb3T again. The locus of this gene was identified by comparing the binary codes

generated by assessing each hybrid as positive or negative for the amplification with the analogous codes for the markers constituting a framework map using the server located at http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl. As a result, BAZ1β was confirmed to be mapped on chromosome 7q11-22 and also located between the markers D7S489 and D7S669 (Fig. 19B).

(4) Analysis of the BAZ1β expression

The cDNA probe of 156 bp prepared by PCR for the testis cDNA using primers nb3S and nb3T was used for Northern analysis of 16 panels of normal tissues (Fig. 20). The probe was labeled with [α-^{32}P]dCTP by random priming and purified with a Chromaspin 10 column (Clontech). Hybridization for Northern analysis was performed at 65°C for 1 hour in ExpressHyb hybridization solution (Clontech). The filters were washed at 65°C until the final stringency reached 0.5 x SSC and 0.1% SDS. Subsequently, autoradiography was performed at -70°C for 4 days to intensify the signals of the filters or for 4 hours with a Fuji BAS system. This probe detected an mRNA of 7.5 kb in almost all the tissues examined. The transcript was analogous to a 7.0 kb transcript of BAZ1α.

Example 5 Expression and purification of BAZ2β fusion protein

Three constructs for BAZ2β were prepared with pGEX vector (Pharmacia) used to express fusion proteins in bacteria. Each of the three constructs contained the sequence corresponding to the amino acid positions 1-190, 1241-1584, or 1500-1970 of BAZ2α (Fig. 21). The expression of the fusion protein was mediated by the IPTG-inducible promoter located upstream from the cloning site. The

expressed proteins were purified through an affinity matrix containing glutathione-Sepharose beads since the expressed protein was fused to glutathione-S-transferase (GST). Specifically, the GST fusion proteins were expressed and purified according to the instructions appended to GST purification modules (Pharmacia). The cultured volume was 400 ml, and proteins were induced by 0.1 mM IPTG at 30°C overnight. Western blotting was performed using BioRad reagents included in an Alkaline Phosphatase Conjugate Substrate kit, according to the manual appended to the kit.

The results of analyzing the expressed protein on the 4-20% gradient SDS-polyacrylamide gel showed that the induced proteins were not detected in the bacterial cell lysates before purification (Fig. 22, Lanes 1, 4, and 5), indicating that the induction through the promoter was not strong in any construct. In any case, however, distinctive proteins (Table 2) with molecular weights corresponding to those predicted were detected (Fig. 23, Lanes 3, 8, and 9). To prove that the purified proteins were the desired fusion proteins, western blot was carried out using the anti-GST antibody. As a result, purified protein with the corresponding size predicted for each protein was detected.

Table 2

Construct	Amino acid region	Predicted MWT kDal	Detected MWT kDal
BAZ2β.1	1-190	51	50
BAZ2β.9	1241-1584	67	65
BAZ2β.11	1500-1970	84	85

Industrial Applicability

The present invention provides a novel transcriptional regulator having a bromodomain, DNA coding said transcriptional regulator, a vector containing said DNA, a transformant expressively retaining said DNA, an antibody binding to said transcriptional regulator, and the method of screening a compound binding to said transcriptional regulator. A transcriptional regulator and DNA of the present invention are expected to be used as indices to diagnose and treat cancer and proliferative diseases, and to screen a drug with a new action mechanism. A compound binding to a transcriptional regulator of the present invention could also be used as a pharmaceutical to treat the diseases described above.

What is claimed is: